The EMBO Journal vol.5 no.5 pp.891–897, 1986

The chicken oestrogen receptor sequence: homology with v-*erbA* and the human oestrogen and glucocorticoid receptors

Andrée Krust, Stephen Green, Patrick Argos[1], Vijay Kumar, Philippe Walter, Jean-Marc Bornert and Pierre Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11, rue Humann, 67085 Strasbourg Cédex, France, and [1]European Molecular Biology Laboratory (EMBL), Postfach 10.2209 Meyerhofstrasse 1, 6900 Heidelberg, FRG

Communicated by P.Chambon

A chicken oviduct cDNA clone containing the complete open reading frame of the oestrogen receptor (ER) has been isolated and sequenced. The mol. wt of the predicted 589-amino acid protein is ~66 kd which is very close to that of the human ER. Comparison of the human and chicken amino acid sequences shows that 80% of their amino acids are identical. There are three highly conserved regions; the second and third of which probably represent the DNA- and hormone-binding domains of the receptor. The putative DNA-binding domain is characterised by its high cysteine and basic amino acid content, and the hormone-binding domain by its overall hydrophobicity. These two domains of homology are also present in the human glucocorticoid receptor (GR) and the product of the avian erythroblastosis virus (AEV) gene, v-*erbA*, indicating that c-*erbA*, the cellular counterpart of v-*erbA*, belongs to a multigene family of transcriptional regulatory proteins which bind steroid-related ligands. The first highly conserved ER region is not present in the truncated v-*erbA* gene, but shares some homology with the N-terminal end of the GR. The function of the v-*erbA* gene product is discussed in relation to its homology with the ER and GR sequences.

Key words: steroid hormones/DNA-binding protein/transcriptional regulation/oncogene/AEV

Introduction

Steroid receptors are eukaryotic transcriptional regulatory factors which, in response to the binding of the steroid hormone, are believed to bind to specific DNA elements and activate transcription (for review, see Yamamoto, 1985). The recent cloning and sequencing of the human oestrogen receptor (ER) (Walter *et al.*, 1985; Green *et al.*, 1986) and human glucocorticoid receptor (GR) (Hollenberg *et al.*, 1985; Weinberger *et al.*, 1985a) has shown that both contain a large degree of homology with the v-*erbA* gene product of the avian erythroblastosis virus (AEV). The AEV genome contains two cell-derived genes termed v-*erbA* and v-*erbB* (for review, see Graf and Beug, 1983). v-*erbA*, expressed as a *gag–erbA* fusion protein, is incapable of erythroblast transformation, but blocks their maturation and potentiates the transforming action of the epidermal growth factor receptor-like oncogene v-*erbB* (Frykberg *et al.*, 1983).

We have now cloned and sequenced the chicken ER cDNA to investigate further the relationship between steroid hormone receptors and *erbA* proteins as well as to determine which regions of the ER are likely to be important for its function, assuming that such regions are highly conserved.

Results

Cloning of the chicken oestrogen receptor

Previous results have shown that the human ER cDNA clone λOR8, which corresponds to the complete open reading frame of the ER (Green *et al.*, 1986), hybridised to a chicken oviduct poly(A)$^+$RNA of ~7.5 kb under stringent hybridisation conditions, indicating a high degree of homology between the two sequences (Walter *et al.*, 1985). Poly(A)$^+$RNA, isolated from laying hen oviduct, was fractionated on denaturing sucrose gradients and aliquots of each fraction analysed on Northern blots using the nick-translated cDNA insert of λOR8. The 7.5-kb RNA, which sedimented faster than 28S, was used to prepare randomly primed cDNA which was cloned into the λgt10 vector after addition of *Eco*RI linkers (Materials and methods). The phage were plated directly, without amplification, and screened in duplicate using the λOR8 cDNA insert as a probe. Sixty three clones were isolated after two rounds of purification. These clones were further analysed using three fragments of the λOR8 cDNA corresponding to the 5′ end, the middle and 3′ end of the ER coding sequence (see Materials and methods). Two clones, λcOR20 (2.1 kb) and λcOR21 (2.0 kb), hybridising with all three probes, were subcloned into pBR322 and mapped. Both clones have the same restriction map and λcOR20 contains all of the λcOR21 sequence. All the other clones hybridise strongly to λcOR20 under stringent conditions, suggesting that they all correspond to partial cDNA sequences of the same RNA. Restriction enzyme fragments of the insert of λcOR20 were subcloned, in both directions, into the M13 mp8 vector and sequenced using the dideoxy method (Figure 1).

The chicken oestrogen receptor sequence

The cOR20 cDNA contains a 589 amino acid long open reading frame (Figure 1) (mol. wt 66 669), which is very similar in size to that of the human ER cDNA [595 amino acids, mol. wt 66 182 (Green *et al.*, 1986)]. There are 202 nucleotides between the initiation codon of this open reading frame and the 5′ extremity of the cOR20 cDNA, indicating that the cER mRNA contains a 5′ leader sequence of at least 202 nucleotides. Two small open reading frames corresponding to 20 (ORF1) and seven (ORF2) amino acids are present in this region. Their significance is unknown, but it is interesting to note that the 232-nucleotide-long 5′ leader sequence of the human ER mRNA also contains a 20-amino acid open reading frame (Green *et al.*, 1986). There is, however, no sequence similarity between the two putative 20-amino acid peptides (Figure 2a). Since the chicken ER mRNA is ~7.5 kb long, it is likely that it contains a very long 3′ untranslated region as is the case for its human counterpart.

The chicken and human oestrogen receptor sequences are closely homologous

The chicken and human ER amino acid sequences are optimally aligned by introducing three gaps of four, two and one amino

acid(s), respectively, in the chicken ER, and a single amino acid gap in the human ER (Figure 2b). The two proteins show an 80% overall identity of their amino acid residues. Three regions of high homology, A (87% amino acid identity), C (100%) and E (94%), are readily discerned. The degree of amino acid identity is lower in the two intervening regions B (56%) and D (38%) and in the carboxy-terminal region F (41%) (see also Figure 3a).

Region C, which is hydrophilic, is unusual in its high content in cysteine, lysine and arginine. Its perfect conservation indicates that it must play a crucial role in ER function. Other studies have shown that it contains sequences which are required for tight binding of the hormone–receptor complex to the nucleus (V.Kumar, S.Green, A.Staub and P.Chambon, in preparation), suggesting that it corresponds to the DNA-binding domain. The conservation of region E indicates that it must also be important for ER function. We have previously suggested that this hydrophobic region contains the oestrogen-binding site (Green et al., 1986). This has been since then directly supported by deletion experiments which have shown that the integrity of the C-terminal half of the human ER is required for oestrogen binding

```
                       EcoRI 1
1     gaattccTCTCCCTTTCCAGTGCTCACCCTGCATTTGTTAGTATACTCCTGCTTGCCGAGATCAGAGGAATACCTATGTATTGCTGCTT
      ORF 1                                                                    ORF 2
83    TAACATGATTTCTGCTGAGCCTCAGAATAGGTTCTGGTGTTTTTTTTAGGTGGACTGCCAGCTGCCGATCTTGCTGTTGACAGTAAAAGCAGATATGTTCCTTGCTCACTGCCATTAGCA

203   ATG ACC ATG ACC CTT CAC ACC AAA GCC TCT GGA GTT ACC CTG CTG CAC CAG ATT CAA GGC ACT GAG CTG GAG ACT CTG AGC AGA CCT CAG
      Met Thr Met Thr Leu His Thr Lys Ala Ser Gly Val Thr Leu Leu His Gln Ile Gln Gly Thr Glu Leu Glu Thr Leu Ser Arg Pro Gln  30

293   CTG AAG ATT CCC CTG GAA CGA TCG CTC AGC GAC ATG TAC GTG GAA AGC AAC AAG ACA GGA GTT TTT AAC TAC CCA GAA GGG GCC ACC TAC
      Leu Lys Ile Pro Leu Glu Arg Ser Leu Ser Asp Met Tyr Val Glu Ser Asn Lys Thr Gly Val Phe Asn Tyr Pro Glu Gly Ala Thr Tyr  60

383   GAT TTT GGT ACT ACC GCT CCA GTG TAC GGC TCT ACT ACA CTC AGT TAT GCC CCT ACT TCT GAA TCT TTT GGG TCA AGC AGC CTG GCA GGA
      Asp Phe Gly Thr Thr Ala Pro Val Tyr Gly Ser Thr Thr Leu Ser Tyr Ala Pro Thr Ser Glu Ser Phe Gly Ser Ser Ser Leu Ala Gly  90

473   TTT CAC TCT CTG AAC AAT GTC CCA CCA AGC CCA GTG GTG TTC TTG CAA ACG GCA CCC CAG CTC TCA CCC TTC ATC CAT CAC CAC AGC CAG
      Phe His Ser Leu Asn Asn Val Pro Pro Ser Pro Val Val Phe Leu Gln Thr Ala Pro Gln Leu Ser Pro Phe Ile His His His Ser Gln 120

563   CAG GTA CCC TAC TAC CTT GAA AAT GAA CAG GGC AGC TTT GGG ATG AGG GAA GCT GCT CCT CCA GCC TTC TAC AGG CCA AGC TCT GAT AAT
      Gln Val Pro Tyr Tyr Leu Glu Asn Glu Gln Gly Ser Phe Gly Met Arg Glu Ala Ala Pro Pro Ala Phe Tyr Arg Pro Ser Ser Asp Asn 150

653   AGG CGC CAC AGC ATT CGT GAG AGG ATG TCT AGT ACC AAT GAG AAA GGG AGC CTG TCC ATG GAG TCC ACC AAG GAG ACC CGG TAC TGT GCT
      Arg Arg His Ser Ile Arg Glu Arg Met Ser Ser Thr Asn Glu Lys Gly Ser Leu Ser Met Glu Ser Thr Lys Glu Thr Arg Tyr Cys Ala 180

743   GTG TGC AAC GAC TAT GCT TCA GGC TAC CAC TAT GGG GTC TGG TCT TGT GAG GGC TGC AAA GCT TTT TTC AAA AGA AGT ATT CAA GGG CAC
      Val Cys Asn Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His 210

833   AAT GAC TAC ATG TGT CCT GCC ACT AAC CAG TGT ACT ATT GAC AAG AAC AGA AGA AAG AGT TGC CAG GCC TGC CGA CTA AGA AAA TGC TAT
      Asn Asp Tyr Met Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr 240

923   GAA GTG GGA ATG ATG AAA GGT GGA ATC CGG AAA GAC CGC AGA GGC GGC GAA ATG ATG AAA CAG AAA CGT CAA AGA GAG GAA CAG GAT TCC
      Glu Val Gly Met Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Glu Met Met Lys Gln Lys Arg Gln Arg Glu Glu Gln Asp Ser 270

1013  AGG AAC GGG GAG GCT TCT TCT ACA GAG CTG CGA GCT CCA ACC CTT TGG ACA AGT CCA CTG GTG GTT AAA CAT AAC AAG AAG AAC AGT CCG
      Arg Asn Gly Glu Ala Ser Ser Thr Glu Leu Arg Ala Pro Thr Leu Trp Thr Ser Pro Leu Val Val Lys His Asn Lys Lys Asn Ser Pro 300

1103  GCT CTG TCT CTG ACA GCA GAA CAG ATG GTC AGT GCC TTG CTG GAA GCT GAG CCA CCT ATA GTT TAT TCT GAA TAT GAC CCC AAT AGA CCA
      Ala Leu Ser Leu Thr Ala Glu Gln Met Val Ser Ala Leu Leu Glu Ala Glu Pro Pro Ile Val Tyr Ser Glu Tyr Asp Pro Asn Arg Pro 330

1193  TTC AAC GAA GCA TCT ATG ATG ACC CTG TTG ACC AAC CTT GCA GAC AGA GAA TTA GTG CAC ATG ATC AAC TGG GCA AAG AGA GTT CCA GGA
      Phe Asn Glu Ala Ser Met Met Thr Leu Leu Thr Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg Val Pro Gly 360

1283  TTT GTG GAT TTA ACA CTC CAT GAT CAG GTC CAT CTG CTG GAA TGT GCC TGG TTA GAG ATA TTG ATG ATC GGC TTA GTC TGG CGC TCC ATG
      Phe Val Asp Leu Thr Leu His Asp Gln Val His Leu Leu Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser Met 390

1373  GAA CAC CCA GGA AAG CTT TTA TTT GCA CCT AAT CTA TTA CTG GAC AGG AAT CAA GGG AAA TGT GTA GAG GGC ATG GTG GAA ATC TTT GAC
      Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp 420

1463  ATG CTA CTG GCT ACT GCT GCT CGG TTT CGG ATG ATG AAC CTT CAA GGG GAG GAA TTT GTG TGC CTT AAG TCC ATC ATC CTG CTC AAT TCT
      Met Leu Leu Ala Thr Ala Ala Arg Phe Arg Met Met Asn Leu Gln Gly Glu Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser 450

1553  GGT GTG TAC ACT TTT CTT TCT AGC ACC TTG AAA TCT CTG GAA GAG AGG GAC TAT ATC CAC CGT GTT CTG GAC AAA ATC ACA GAT ACT CTG
      Gly Val Tyr Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Arg Asp Tyr Ile His Arg Val Leu Asp Lys Ile Thr Asp Thr Leu 480

1643  ATA CAC CTA ATG GCA AAG TCA GGT CTT TCT CTG CAG CAG CAA CAC CGG CGA CTA GCT CAG CTC CTC CTT ATC CTC TCT CAC ATC AGG CAT
      Ile His Leu Met Ala Lys Ser Gly Leu Ser Leu Gln Gln Gln His Arg Arg Leu Ala Gln Leu Leu Leu Ile Leu Ser His Ile Arg His 510

1733  ATG AGC AAC AAA GGA ATG GAG CAC CTG TAC AAT ATG AAG TGT AAA AAT GTA GTT CCG CTC TAC GAC CTC TTA CTG GAG ATG CTG GAC GCT
      Met Ser Asn Lys Gly Met Glu His Leu Tyr Asn Met Lys Cys Lys Asn Val Val Pro Leu Tyr Asp Leu Leu Leu Glu Met Leu Asp Ala 540

1823  CAC CGC CTA CAT GCA CCG GCA GCC AGG AGT GCT GCA CCA ATG GAA GAG GAG AAC CGA AAC CAA CTG ACA ACC GCA CCA GCT TCA TCT CAT
      His Arg Leu His Ala Pro Ala Ala Arg Ser Ala Ala Pro Met Glu Glu Glu Asn Arg Asn Gln Leu Thr Thr Ala Pro Ala Ser Ser His 570

1913  TCC CTG CAG TCC TTT TAC ATT AAC AGC AAA GAA GAG GAG AGT ATG CAG AAT ACA ATA TAA ACA AAA GTC AGC ATA TAA TAA TGG TAT GAG
      Ser Leu Gln Ser Phe Tyr Ile Asn Ser Lys Glu Glu Glu Ser Met Gln Asn Thr Ile ***                                     589

2003  AAACCCAGCAGCGTACTACCTAGCTCATGCTTCATTTCATCTATAGTGCCACATATGTCAACACTCCAATGACATTAGCCACCggaattc...
                                                                                         EcoRI
```

Fig. 1. Coding sequence of the chicken oestrogen receptor. The nucleotide sequence of the chicken ER clone λcOR20 is shown from the first nucleotide after the *Eco*RI linker (GGAATTCC) to nucleotide 2088. The numbers on the left refer to the position of the nucleotides and on the right to that of the amino acids. The sequence contains three open reading frames of 20 (ORF 1), seven (ORF 2) and 589 amino acids, respectively. The initiation and stop codon for each of these reading frames is boxed.

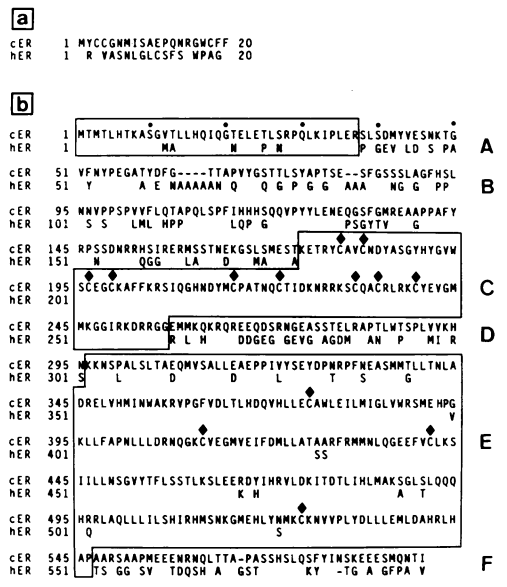

Fig. 2. Comparison of the chicken and human oestrogen receptor amino acid sequences. (**a**) Comparison of the 20 amino acid open reading frame of both the chicken (cER) and human (hER) ERs. Only non-identical amino acids are shown in the hER sequence. (**b**) Comparison of the cER (589 amino acids) and hER (595 amino acids) sequences. Only non-identical amino acids are shown in the hER sequence. Gaps were introduced into the sequence (−) in order to obtain maximal alignment of identical amino acids. The numbers on the left hand side refer to the amino acid sequence. The sequence was divided into six regions (A−F) based on sequence homology between cER and hER. The three most highly conserved regions [A, amino acids 1−37; C, 180−262; E, 302−552 (amino acid numbers with respect to hER)] are boxed. Conserved cysteines (◆) are indicated.

(V.Kumar, S.Green, A.Staub and P.Chambon, in preparation). The high degree of homology present in region A suggests that it is an important functional domain of the ER, whereas the lower conservation of regions B, D and F suggests that their integrity is less important for ER function.

Domain conservation in the oestrogen and glucocorticoid receptors and v-erbA

Sequence homologies have been found between the AEV oncogene potentiator gene v-*erbA* and both the human glucocorticoid (Weinberger *et al.*, 1985b) and oestrogen (Green *et al.*, 1986) receptors. These three sequences, together with that of the chicken ER, have been computer-aligned for maximal homology (see Argos, 1985 for details). As shown in Figures 3 and 4, homologies are revealed between the four sequences by the introduction of few amino acid gaps. Starting at the N-terminus of v-*erbA*, the mean correlation coefficient, using six amino acid physical characteristics (Argos *et al.*, 1983), between the sequences of hER and hGR, hER and v-*erbA* and hGR and v-*erbA* are 0.39, 0.35 and 0.38, respectively (these values are 9.1, 8.2 and 8.6 SD above the control mean, respectively).

Both hGR and v-*erbA* contain two regions which are significantly homologous to the two highly conserved regions C and E of the ERs (Figures 3 and 4). It is striking that all nine cysteines present in region C of the ERs are conserved in both GR and v-*erbA* (Figures 3e and 4). Similarly 8/10 basic amino acids present in the ER are fully conserved, whereas the remaining two, whilst identical in the GR, represent conservative changes as DNA-binding amino acids in v-*erbA* (Arg → Thr, Lys → Asn; Ohlendorf and Matthews, 1983). The homology in region E, corresponding to amino acids 315−540 of the hER

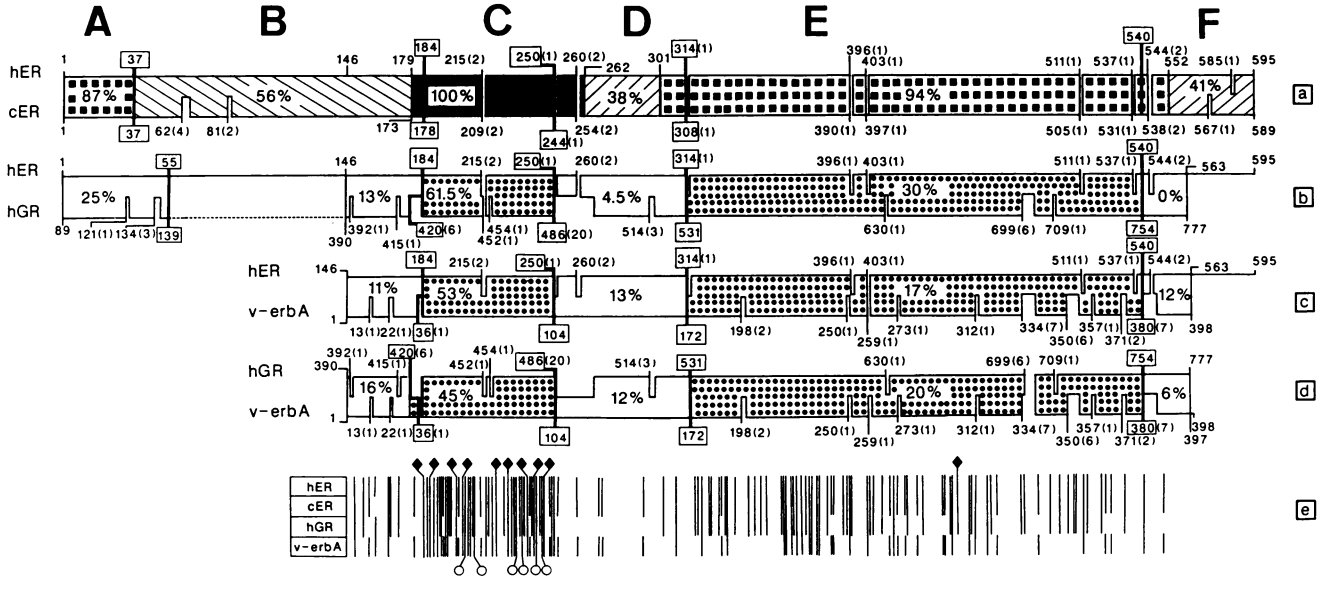

Fig. 3. Schematic alignment of the oestrogen and glucocorticoid receptors and the v-*erbA* protein. The four amino acid sequences (hER, human oestrogen receptor; cER, chicken oestrogen receptor; hGR, human glucocorticoid receptor; and v-*erbA*) were aligned by computer introducing gaps for maximal alignment of all four sequences using the criteria as described by Argos (1985). The alignment shown here is identical to that shown in Figure 4. The positions of gaps introduced into the sequences to maximise the alignment are shown with the length of each gap given in parentheses. The sequences were then compared in pairs: (**a**) hER with cER, (**b**) hER with hGR, (**c**)hER with v-*erbA* and (**d**) hGR with v-*erbA*. The hER/cER aligned sequences have been divided into six regions (A−F) based on their homology (see also Figure 2). The position of the amino acids corresponding to the boundaries of these regions which are common to all four sequences are boxed. No homology was observed between residues 140 and 389 of the hGR sequence and hER (dotted line). Shaded regions refer to compared sequences possessing significant homology (greater than 3 standard deviations above the control mean, see Figure 4). The correlation of the alignment (see Figure 4) between two sequences was calculated over six amino acid physical characteristics (see Argos *et al.*, 1983; Argos, 1985) and is dependent upon both the similarity and length of the aligned sequences. The percentages of aligned amino acid residues which are identical in each region are indicated. Those amino acids which are identical in at least three out of four of the sequences are indicated in e (see also Figure 4). Highly conserved cysteines (◆) and basic amino acids (○) are indicated (see also Figure 4).

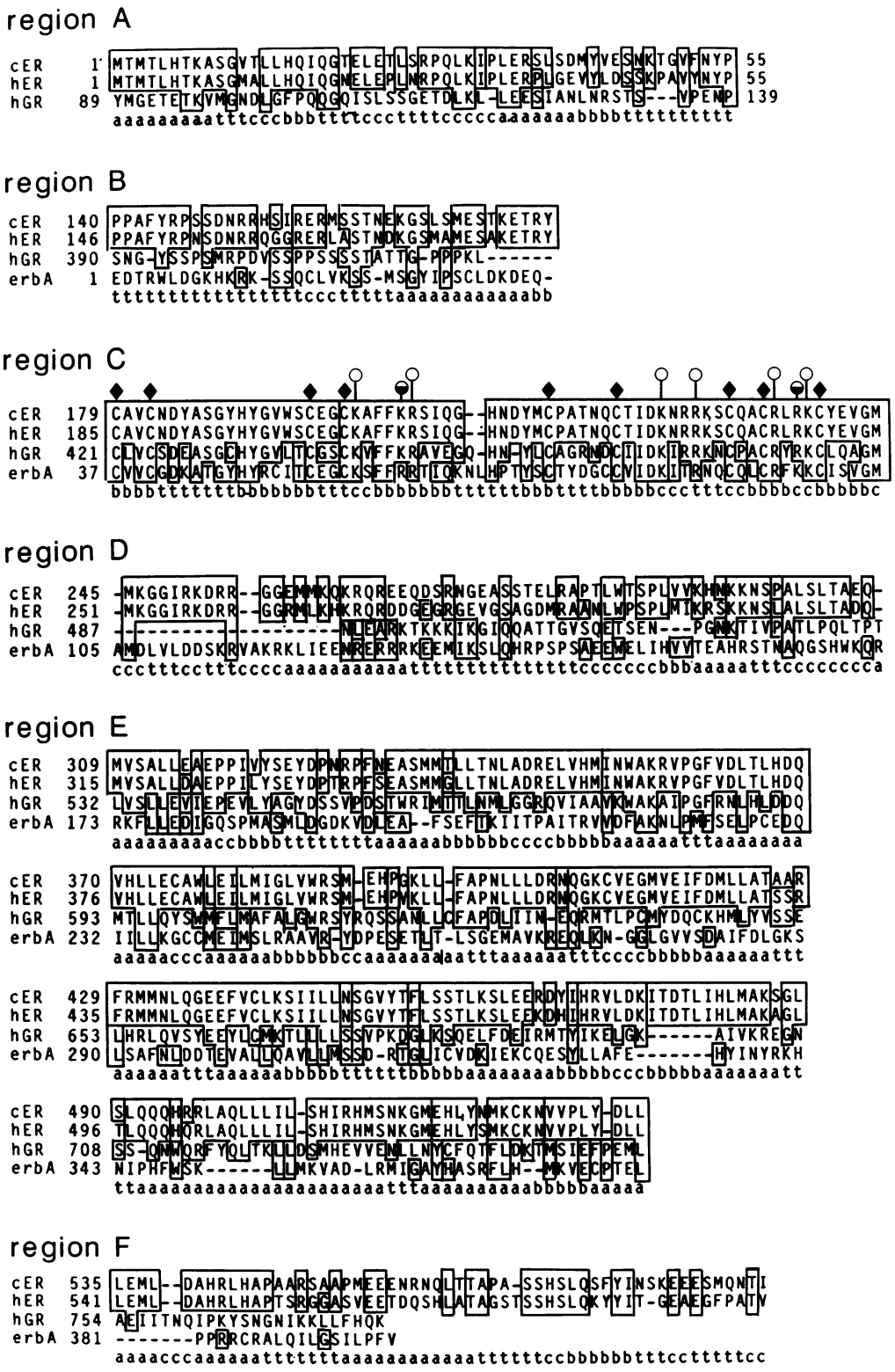

Fig. 4. Alignment of the oestrogen and glucocorticoid receptors and the v-*erbA* protein. The four sequences, cER, hER, hGR and v-*erbA*, were aligned as described in Figure 3. The aligned sequences have been divided into six regions (A−F) based on their shared homology. Note that the last 11−14 C-terminal amino acids of the v-*erbA* sequence appear to be derived from *env*-related and c-*erbB* intron sequences (Henry *et al.*, 1985). Numbers represent the position of the amino acid residues in each sequence. Identical amino acids, in two or more sequences, are boxed. The predicted secondary structure, based on the average of the aligned residues (see Argos, 1985 for details), is indicated below the lower line of each alignment (a = helix, b = β strand, t = turn, c = coil). In region C: ♦, represents fully conserved cysteine residues; ○, fully conserved lysine or arginine residues; and ◒, fully conserved basic residues (mixture of lysine or arginine). The mean correlation coefficients and number of standard deviations above the control mean (in parentheses) (see Argos, 1985) for the hER/hGR, hER/v-*erbA* and hGR/v-*erbA* alignments for each of the different regions are as follows: A, (hER/hGR only) 0.30, (2.5σ); B, −0.05, 0.15, 0.32 (−, 1.1σ, 2.3σ); C, 0.68, 0.63, 0.64 (6.4σ, 5.9σ, 6.0σ); D, 0.0, 0.24, 0.27 (−, 2.2σ, 2.3σ); E, 0.41, 0.33, 0.35, (7.3σ, 5.6σ, 5.9σ); F, 0.22, 0.03, 0.27 (2.2σ, −, 2.5σ). Values for the overall alignments (B−F) are: 0.39 (9.1σ) for hER/hGR; 0.35 (8.2σ) for hER/v-*erbA*; 0.38 (8.6σ for hGR/v-*erbA*. (For details of the statistical procedure see Argos, 1985; Green *et al.*, 1986.)

sequence (Figures 3 and 4), is greater between the hER and the hGR than between either of them and v-*erbA*. The average correlation coefficients are 0.41 (hER/hGR), 0.33 (hER/v-*erbA*) and 0.35 (hGR/v-*erbA*), with standard deviations above the control mean of 7.3, 5.6 and 5.9, respectively. The conservation of quite a number of amino acids in all four sequences suggests that the overall structure of this hydrophobic region may be similar in all four proteins. Note, however, that none of the positions of the cysteine residues are fully conserved. Region D, which separates the conserved domains C and E, is not significantly conserved between the ER, GR and v-*erbA* sequences and is markedly shorter in the GR sequence (Figures 3 and 4). No significant homology could be detected between region B of the ER and the corresponding region of the GR and v-*erbA* (Figures 3 and 4). There is, however, some homology between region A of the ER and a portion of the N-terminal region of the GR (Figures 3 and 4).

Discussion

Comparison of the chicken and human ER amino acid sequences shows the presence of three highly conserved regions A, C and E (Figures 3a and 4). Extending this comparison to the human GR and v-*erbA* sequences shows that homologous counterparts of these three regions are similarly arranged in the hGR, whereas v-*erbA* contains sequences homologous to regions C and E only (Figures 3 and 4). This leaves little doubt that these two steroid hormone receptors and the *erbA* genes are all derived from a common primordial gene.

The hydrophilic region C, which is rich in arginine and lysine, is remarkable in that the position of nine cysteine residues is fully conserved (9/9, 9/10 and 9/11 of the cysteines of the ERs, GR and v-*erbA*, respectively). As mentioned above (V.Kumar *et al.*, in preparation), sequences present in this region are required to ensure tight association of the oestrogen receptor complex to the nucleus when the ER is transiently expressed in HeLa cells (Green *et al.*, 1986). It has been proposed that steroid–hormone receptor complexes exert their transcriptional regulatory function by binding to specific promoter elements (for review see Yamamoto, 1985). This proposal has been experimentally supported in the case of the glucocorticoid and progesterone receptors (for references and review, see Von der Ahe *et al.*, 1985; Yamamoto, 1985). Thus, it is most probable that region C corresponds to a DNA-binding domain in which the conserved cysteine residues may constitute the common structural scaffolding and/or act to facilitate signal transmission after the receptor binds its ligand, whereas the specificity of DNA-binding could reside in some of the less conserved amino acids. DNA-binding domains of prokaryotic transcriptional regulatory proteins contain an α-helix–β-turn–α-helix secondary structure, in which the two helices provide the binding specificity through intimate contact in the major groove of the B form of the DNA helix (for references see Pabo and Sauer, 1984; Anderson *et al.*, 1985). The *Xenopus* 5S transcription factor, TFIIIA (Brown *et al.*, 1985; Miller *et al.*, 1985), and the product of a *Drosophila* segmentation gene, *Krüppel* (Rosenberg *et al.*, 1986), contain repeated motifs rich in cysteine, histidine and basic amino acid residues, which are thought to bind Zn^{2+} and may correspond to another kind of DNA-binding domain (for review and additional references, see Berg, 1986). The putative DNA-binding domain described here is unlikely to form a helix–turn–helix secondary structure especially given the high β-strand–turn prediction (Figure 4) and the arrangement of its cysteine and histidine residues is markedly different from the TFIIIA motif. It may therefore be characteristic of a new family of regulatory DNA-binding proteins.

The hydrophobic region E, which is highly conserved in the human and chicken ER, is less well conserved in hGR and v-*erbA* than is region C. As mentioned, deletions within domain E have shown that its integrity is required for the binding of oestradiol by the hER using a HeLa cell transient transfection assay (Green *et al.*, 1986). Sequences present in the carboxy-terminal region of the hGR appear to be also required for hormone binding (Hollenberg *et al.*, 1985). The affinity of steroid hormones for the various receptors can differ considerably. Thus, the moderate conservation of the putative steroid hormone-binding domain E between the ER and GR may reflect a common overall structure, creating a hydrophobic pocket which contains the steroid-binding site whose specificity would reside in some of the non-conserved amino acids. The secondary structure prediction for this region suggests a domain primarily consisting of α-helices and β-strand, a structural mix typical in known tertiary structures which form hydrophobic pockets (e.g. NAD-binding domains and dehydrogenases; Rossman *et al.*, 1974). The lack of positional conservation of the cysteine residues between the ERs, GR and v-*erbA* in domain E suggests that either cysteine bridges are unimportant for the shape of such a pocket or that its shape is different between the ER and GR. Comparison of the four proteins suggests that the common domain E has a minimum size of 226 amino acids (315–540 in the hER) with a mol. wt of ~25 kd. It is interesting that proteolytic cleavage of the rat GR has localised the hormone-binding site on a 25-kd peptide fragment containing no DNA-binding activity (Wrange *et al.*, 1984). This is compatible with trypsin cleavage within region D, which is rich in basic amino acids and likely to be exposed.

Upon binding their cognate steroid hormones, the receptors become more tightly bound to the nucleus, presumably because their affinity for specific DNA sequences is increased (for references, see Yamamoto *et al.*, 1985). The underlying mechanism is unknown, but may involve communication between the hormone-binding and DNA-binding domains. That region D, which separates domains C and E, is the least conserved when comparing the two ERs and is not significantly conserved (in neither length nor sequence) when all four proteins are compared, suggests that its function does not require a highly specific structure. The predicted secondary structure of the highly hydrophilic region D is rich in turn and coil and thus likely to be flexible. It is more probable, therefore, that communication between domains C and E involves direct contacts between them, rather than the transmission of a signal through region D; the latter may provide a hinge between the two domains. Such a structural arrangement containing two important functional domains separated by a hinge has been recently described for prokaryotic and eukaryotic DNA-binding transcriptional regulatory proteins (see references in Brent and Ptashne, 1985). Comparison of all four protein sequences (Figures 3 and 4) suggests a minimum size of 356 amino acids (185–540 in the hER; mol. wt ~39 kd) for the receptor region containing both the DNA- and the hormone-binding domains. This is in agreement with the results of chymotrypsin cleavage studies with the rat GR which have shown that a 39-kd fragment contains both DNA- and steroid hormone-binding domains (Wrange *et al.*, 1984).

The conservation of region A between the human and chicken ERs suggests that it plays an important role in the function of the ER. The mechanism of ER action may require the receptor to interact with some element(s) of the transcription machinery in addition to its interactions with the hormone and the DNA.

The N-terminal region A may perform such a function. In this respect, we note that this region appears also to be partially conserved at the amino acid end of the GR. Whether a similar region is also present in the N-terminal region of the cellular counterpart (c-*erbA*) of v-*erbA* will require its cloning and sequence determination, since v-*erbA* corresponds to a truncated c-*erbA* fused to the AEV *gag* gene (Debuire *et al.*, 1984).

Region B, which is moderately conserved in the human and chicken ERs, is not conserved at all in the hGR, where it contains the immunogenic domain whose function is unknown. Weinberger *et al.* (1985b) have found a limited match between short sequences present in this region and the sequences of the homeotic proteins encoded by the *Drosophila Antennapedia* genes (Scott and Weiner, 1984) and *Fushi tarazu* (Laughton and Scott, 1984). No such homologies could be detected between these genes and the ERs. Steroid hormone receptors are often associated with 90-kd heat-shock protein(s) which are present in cells in large amounts (see Catelli *et al.*, 1985 for references). Since the receptor domain which is recognized by this protein should be common to both the ER and GR, it is unlikely that it could be either of the two less conserved regions B and F, unless the interaction region is small and weakly conserved.

v-*erbA* is expressed as the p75 *gag−v−erbA* fusion protein in AEV-transformed cells. It has no transforming activity by itself but potentiates the transforming activity of v-*erbB* and appears to be responsible for the early blockage of erythroid cell differentiation (Frykberg *et al.*, 1983). The function of c-*erbA*, the cellular counterpart of v-*erbA*, as well as that of a related gene which has been found in the human genome (Jansson *et al.*, 1983), is unknown. That v-*erbA* contains sequences homologous to both the putative DNA- and hormone-binding domains of the ERs, which are also conserved in the hGR, strongly supports the view that c-*erbA* belongs to a multigene family of DNA-binding transcriptional regulatory proteins, which bind steroid-related molecules and are all derived from a common ancestor. Assuming that v-*erbA* can exert the same function as c-*erbA*, the blockage of erythroblast maturation at an early stage of differentiation may be due to the inappropriate expression of c-*erbA*-controlled genes, which are not normally expressed at this stage of differentiation. On the other hand, assuming that the structure of the c-*erbA* protein is similar to that of the oestrogen and glucocorticoid receptors, v-*erbA*, which is a truncated c-*erbA*, may contain the DNA- and hormone-binding domains C and E, but not the N-terminal regions A and B. This may allow the v-*erbA* protein to bind to c-*erbA* DNA regulatory sites, but, due to the absence of regions A and B, it may be unable to activate transcription. This possibility raises the interesting alternative explanation that the v-*erbA* protein could block erythroblast differentiation by competing for the same DNA regulatory sites as c-*erbA*. The normal function of c-*erbA* should then be to activate the transcription of genes whose expression is required at an early stage of differentiation for erythroblast maturation. In support of such a model, it has been shown that a mutant of the yeast *GAL4* transcriptional regulatory protein, containing the DNA-binding domain but lacking transcriptional activator domains, is able to repress those same genes which it normally activates (Keegan *et al.*, 1986).

Further characterisation of c-*erbA*, site-directed mutagenesis experiments and the creation of chimeric proteins by swapping over the various domains of the ER, GR and c-*erbA* proteins will be helpful in elucidating the function of the various steroid hormone receptor domains, in characterising the relationship between the various members of this new multiple family and in explaining how some of them can potentiate the transforming activity of oncogenes.

Materials and methods

Sucrose gradient enrichment of chicken ER mRNA

Total RNA was isolated from laying hens (Auffray and Rougeon, 1980) and purified by oligo(dT)-cellulose chromatography. One hundred microgrammes of this RNA were fractionated on 5−20% methyl mercury hydroxide-containing sucrose gradients (Walter *et al.*, 1985). One-tenth of each of the 28 fractions was fractionated on a 1% agarose gel containing 10 mM methyl mercury hydroxide (Bailey and Davidson, 1976), transferred to diazobenzyl-oxymethyl (DBM) paper (Alwine *et al.*, 1979) and hybridised with the nick-translated insert of λOR8 (Walter *et al.*, 1985).

Cloning of the chicken ER cDNA in λgt10

Those fractions of the sucrose gradient which were enriched in ER mRNA were pooled and used to prepare randomly primed cDNA, suitable for insertion into the λgt10 vector (Huynh *et al.*, 1985), as previously described (Walter *et al.*, 1985). The phage were plated onto *Escherichia coli* C600 hfl, without amplification, with ~3000−5000 phage per 8.5 cm Petri dish. Duplicate filters were screened with the nick-translated insert of λOR8 (sp. act. ~2 × 10^8 c.p.m./μg. 200 000 c.p.m. per filter) overnight at 65°C (0.6 M NaCl, 0.06 M Tris-HCl pH 7.5, 0.004 M EDTA, 0.1% polyvinylpyrolidone, 0.1% bovine serum albumin, 0.1% ficol, 0.2% SDS, 50 μg/ml salmon sperm DNA). The filters were washed three times for 30 min at 65°C in hybridisation buffer without salmon sperm DNA, followed by a 30 min wash at room temperature in 2 X SSC (0.3 M NaCl, 0.03 M sodium citrate). Sixty-three clones, positive after two rounds of screening, were isolated from 860 000 phages. In order to isolate a chicken clone containing the complete ER coding sequence, these clones were screened using three subclones of λOR8 corresponding to the 5′ end [*Sma*I(77)−*Sma*I(476)], the middle [*Sma*I(476)−*Hin*dIII(1248)] and 3′ end [*Hin*dIII(1248)−*Eco*RI(2030)] of the ER coding sequence (Green *et al.*, 1986).

Sequencing

The 2.1-kb insert of λcOR20 was subcloned into the *Eco*RI site of pBR322, mapped, and further subcloned into M13mp8. Sequencing was performed using the dideoxy technique on both strands of the DNA. Sequencing was primed using either the M13 universal primer or synthetic oligonucleotide primers corresponding to the chicken ER cDNA sequence.

Acknowledgements

We thank A.Staub for synthesis of the oligonucleotides and I.Issemann for her excellent technical assistance. We would also particularly like to thank C.Werlé and B.Boulay for the preparation of the figures. This research was supported by the INSERM, the CNRS (ATP Etats-Uni), and the Ministère de la Recherche et de la Technologie (84V0803). Stephen Green is a recipient of a Royal Society European Exchange fellowship.

References

Alwine,J.C., Kemp,D.J., Parker,B.A., Reiser,J., Renart,J., Stark,G.R. and Wahl,G.M. (1979) *Methods Enzymol.*, **68**, 220−242.
Anderson,J.E., Ptashne,M. and Harrison,S.C. (1985) *Nature*, **316**, 596−601.
Argos,P., Hanei,M., Wilson,J.M. and Kelley,W.N. (1983) *J. Biol. Chem.*, **258**, 6450−6457.
Argos,P. (1985) *EMBO J.*, **4**, 1351−1355.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.* **107**, 303−314.
Bailey,J.M. and Davidson,N. (1976) *Anal. Biochem.*, **70**, 75−85.
Berg,J.M. (1986) *Nature*, **319**, 264−265.
Brent,R. and Ptashne,M. (1985) *Cell*, **43**, 729−736.
Brown,R.S., Sander,C. and Argos,P. (1985) *FEBS Lett.*, **186**, 271−274.
Catelli,M.G., Binart,N., Jung-Testas,I., Renoir,J.M., Baulieu,E.E., Feramisco,J.R. and Welch,W.J. (1985) *EMBO J.*, **4**, 3131−3135.
Debuire,B., Henry,C., Benaissa,M., Biserte,G., Claverie,J.M., Saule,S., Martin,P. and Stehelin,D. (1984) *Science*, **224**, 1456−1459.
Frykberg,L., Palmieri,S., Beug,H., Graf,T., Hayman,M.J. and Vennström,B. (1983) *Cell*, **32**, 227−238.
Graf,T. and Beug,H. (1983) *Cell*, **34**, 7−9.
Green,S., Walter,P., Kumar,V., Krust,A., Bornert,J.M., Argos,P. and Chambon,P. (1986) *Nature*, **320**, 134−139.
Henry,C., Coquillaud,M., Saule,S., Stehelin,D. and Debuire,B. (1985) *Virology*, **140**, 179−182.
Hollenberg,S.M., Weinberger,C., Ong,E.S., Cerelli,G., Oro,A., Lebo,R., Thompson,E.B., Rosenfeld,M.G. and Evans,R.M. (1985) *Nature*, **318**, 635−641.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning — A Practical Approach*. IRL Press, Oxford; Vol. 1, pp. 49–78.
Jansson,M., Philipson,L. and Vennström,B. (1983) *EMBO J.*, **2**, 561–565.
Keegan,L., Gill,G. and Ptashne,M. (1986) *Science*, in press.
Laughton,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Ohlendorf,D.H. and Matthews,B.W. (1983) *Annu. Rev. Biophys. Bioeng.*, **12**, 259–284.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.
Rosenberg,U.B., Schröder,C., Preiss,A., Kienlin,A., Côté,S., Reide,I. and Jäckle,H. (1986) *Nature*, **319**, 336–339.
Rossman,M.G., Moras,D. and Olsen,K.W. (1974) *Nature*, **250**, 194–199.
Scott,M.P. and Weiner,A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115–4119.
Von der Ahe,D., Janich,S., Scheidereit,C. and Beato,M. (1985) *Nature*, **313**, 705–709.
Walter,P., Green,S., Greene,G.L., Krust,A., Bornert,J.M., Jeltsch,J.M., Staub,A., Jensen,E., Scrace,G., Waterfield,M. and Chambon,P. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7889–7893.
Weinberger,C., Hollenberg,S.M., Ong,E.S., Harmon,J.M., Brower,S.T., Cidlowski,J., Thompson,E.B., Rosenfeld,M.G. and Evans,R.M. (1985a) *Science*, **228**, 740–742.
Weinberger,C., Hollenberg,S.M., Rosenfeld,M.G. and Evans,R.M. (1985b) *Nature*, **318**, 670–672.
Wrange,O., Okret,S., Radofcic,M., Carlstedt-Duke,J. and Gustafsson,J.A. (1984) *J.Biol. Chem.*, **259**, 4534–4541.
Yamamoto,K.R. (1985) *Annu. Rev. Genet.*, **19**, 209–252.

Received on 3 February 1986; revised on 17 February 1986

Note added in Proof

Hartshorne *et al.* [Hartshorne,T.A., Blumberg,H. and Young,E.T. (1986) *Nature*, **320**, 283–287], have recently pointed out that two yeast transcriptional regulatory proteins, GAL4 [Laughon,A. and Gesteland,R.F. (1984) *Mol. Cell. Biol.*, **186**, 260–267] and PPR1 [Kammerer,B., Guyonvarch,A. and Hubert,J.C. (1984) *J. Mol. Biol.*, **180**, 239–250] contain a putative DNA-binding domain which shares the cysteine-rich motif, Cys-X-X-Cys-$(X)_{13}$-Cys-X-X-Cys, which is also present in steroid hormone receptors begining at Cys^{185} in the hER (see Figure 4). These authors suggest that such motifs, with two pairs of cysteine residues, may also form a DNA-binding 'finger' with the four cysteine residues bound to Zn^{2+} and some of the other amino acids making contact with the DNA. The C-terminal half of the putative DNA-binding domain of the steroid hormone receptors (216–250 for hER) does not contain the same motif, but is rich in cysteine, histidine and basic amino acids and may, therefore, be also organized into some kind of DNA-binding 'finger'.